STOCHASTIC MODELS IN A FREE-RECALL EXPERIMENT

by

DAVID LOYAL JENKINS

B. S., Fort Hays Kansas State College, 1964

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics and
Statistical Laboratory

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1967

Approved by

W J Conover

Major Professor

CONTENTS

STOCHASTIC MODELS IN A FREE-RECALL EXPERIMENT

1. Introduction

The purpose of this paper is to show some of the stochastic models used to represent the data of a free-recall experiment often done in psychology. The use of stochastic models is relatively new. Most of the work has been done since 1950.

This paper will consider four stochastic models. Choice of a model will depend on assumptions and experimental procedures. It is a well known fact in psychology that two experiments on the same variable may not yield the same results and conclusions. Two experimenters studying the same variable may vary the other variables in their experiment differently. So before a stochastic model can be used the experimental procedure must be specified. The free recall procedure for the first three models considered in this paper the subject is allowed as many trials as he needs to learn completely the list of words. In the last model considered the subject may only partially learn the list of words.

A stochastic model provides a framework for analyzing data at the level of single subjects and single trials. Models also provide a way to summarize the data once various parameters are estimated. For a general discussion of the value of stochastic models see Miller (1964).

The first model considered is a model by Bush and Mosteller (1955). They assumed that if a word is recalled on a trial the probability of recalling the word on the next trial is increased, and this change could be represented by linear operators.

The second model is the Miller and McGill (1952) model. Their model is closely related to Bush and Mosteller's model. Use of some results from Bush and Mosteller's model in Miller and McGill's model yields several interesting results.

The third model considered is a model by Waugh and Smith (1962). Their model is a Markov chain with an absorption state. They look at the words in terms of what state they are in.

The last model considered is the model by Cowan (1966). This model is not like the previous models. Cowan's model considers the fact that certain words tend to cluster together in recall.

2. Description of a Free-Recall Experiment

A list of nonsense syllables or monosyllabic words is presented to a subject. At the end of the presentation the subject writes down all the words that he can recall. The order of the words is then randomized, and the procedure is repeated until the list is completely learned.

3. The Bush-Mosteller Linear Model

In developing their model Bush and Mosteller were influenced by a paper written by Estes (1950). The authors first

described in two papers, (1951) and (1953), the basic structure of the linear model. Since then they have published many other articles and books on their model, most of which are listed in Atkinson, Bower, and Crothers (1965).

3.1 Definitions and Terms.

Let p be the probability of a word being recalled, and q be the probability of that word not being recalled. These are two mutually exclusive events. Either the word is recalled, denoted by E_1, or the word is not recalled, denoted by E_2. It is assumed that whenever either of the two events occurs the probabilities of recall or non-recall are altered. So, corresponding to each event there is a mathematical operator T_i $(i=1,2)$ which when applied to the probabilities, transforms the probabilities to the probabilities of recall or non-recall on the succeeding trial. Bush and Mosteller (1951) considered the case where the operation Op was expressible as a power series in p. They considered the function $Tp=a_0+a_1p$ as an approximation to the function Op. Since Tp was a linear function of p, then matrix operators could be used.

3.2 The General Model.

Bush and Mosteller (1955) considered that the event E_i had a matrix operator T_i of the general form

$$T_i = \begin{bmatrix} u_{11,i} & u_{12,i} \\ u_{21,i} & u_{22,i} \end{bmatrix} , \quad i = 1, 2 \ .$$

Applying the operator T_i to the probability vector $\vec{p}' = (p,q)$, the vector $T_i p$ is obtained.

$$T_i p = \begin{bmatrix} u_{11,i}p + u_{12,i}q \\ u_{21,i}p + u_{22,i}q \end{bmatrix}$$

The probability of recalling a word on the next trial after event E_i occurs is $u_{11,i}p + u_{12,i}q$, whereas the probability of non-recall is $u_{21,i}p + u_{22,i}q$. These new probabilities must sum to one. So

$$(u_{11,i}p + u_{12,i}q) + (u_{21,i}p + u_{22,i}q) = 1$$

or

$$(u_{11,i} + u_{21,i})p + (u_{12,i} + u_{22,i})q = 1 \quad .$$

The above equation must hold for all values of p and q consistent with the condition that p and q sum to unity, and so in particular for $p = 1$ and $q = 0$,

$$u_{11,i} + u_{21,i} = 1$$

whereas for $q = 1$ and $p = 0$

$$u_{12,i} + u_{22,i} = 1 \quad .$$

These equations mean the columns of the matrix T_i must sum to unity. Letting $a_i = u_{12,i}$ and $b_i = u_{21,i}$ the matrix operator T_i may be written as

$$T_i = \begin{bmatrix} 1-b_i & a_i \\ b_i & 1-a_i \end{bmatrix} .$$

Applying the operator T_i to the probability vector $\vec{p}$ gives

$$T_i\vec{p} = \begin{bmatrix} (1-b_i)p+a_iq \\ b_ip+(1-a_i)q \end{bmatrix} , \quad i = 1, 2 .$$

Let Q_ip and $\tilde{Q}_iq$ denote the first and second element of vector $T_i\vec{p}$ respectively. Letting $\alpha_i = 1-a_i-b_i$, $a_i = (1-\alpha_i)\lambda_i$, and using the fact that $p = 1-q$ the element Q_ip may be written as

(3.2.1) $\qquad Q_ip = \alpha_ip + (1-\alpha_i)\lambda_i \quad .$

Bush and Mosteller (1955) have shown that for Q_ip to be between the limits of zero and one and to represent learning probabilities, then $0 \leq \alpha_i \leq 1$ and $0 \leq \lambda_i \leq 1$ must hold. Note, that Q_ip is the probability of recalling a word on the next trial after event E_i has occurred.

On succeeding trials either E_1 or E_2 occurs. The occurrence of E_i means that the operator Q_i must be applied to the probability Q_ip.

$$\begin{aligned} Q_i(Q_ip) &= \alpha_i(Q_ip) + (1-\alpha_i)\lambda_i \\ &= \alpha_i(\alpha_ip + (1-\alpha_i)\lambda_i) + (1-\alpha_i)\lambda_i \\ &= \alpha_i^2p + (1-\alpha_i^2)\lambda_i \quad . \end{aligned}$$

The forms of Q_i^2p and Q_ip suggests the general form for any number n of applications is

$$Q_i^n p = \alpha_i^n p + (1-\alpha_i^n)\lambda_i \quad .$$

Using mathematical induction, the general form can be proven to be true. Now, when α_i is less than unity, α_i^n tends to zero as n gets large, so

(3.2.2) $$\lim_{n\to\infty} Q_i^n p = \lambda_i \quad .$$

3.3 Assumptions Made for Free-Recall Experiments.

To simplify the estimation problem of the parameters Bush and Mosteller (1955) made certain assumptions. The first assumption made was that the probability of recalling one word is independent of the other words. The second assumption made was that all words have the same initial probability of recall, p_0. The third assumption was that all words were equally difficult to learn and the position on the list was irrelevant. The fourth assumption made was that the non-recall of a word doesn't change its probability of being recalled on the next trial. The fifth assumption made was that a subject could learn a list of words perfectly.

3.4 What the Assumptions Mean to the Model.

Let the probability that the i^{th} word is recalled on trial n be $p_{i,n}$. Now given that the i^{th} word is not recalled on the n^{th} trial the probability of recall on the (n+1)th trial is not changed, by the fourth assumption. So Q_2, which is applied when E_2 occurs, must be the identity operator. This means that (3.2.1) becomes

(3.4.1) $$p_{i,n+1} = Q_2 p_{i,n} = p_{i,n} \quad .$$

For this equation to be of this form, then α_2 must be equal to one. Using the last assumption and (3.2.2), then $\lambda_1 = 1$. This means that

(3.4.2) $$p_{i,n+1} = Q_1 p_{i,n} = \alpha_1 p_{i,n} + (1-\alpha_1) \quad .$$

Since all the words start with the same initial probability of recall, then any words that have been recalled exactly k times will have the same probability p_k of recall on the next trial. To find the probability of recalling a word after k recalls, the operator Q_1 would be applied k times. The first application yields

$$p_1 = Q_1 p_0 = \alpha_1 p_0 + (1-\alpha_1) \quad .$$

The probability after two recalls is

$$\begin{aligned} p_2 &= Q_1(Q_1 p_0) = \alpha_1(Q_1 p_0) + (1-\alpha_1) \\ &= \alpha_1[\alpha_1 p_0 + (1-\alpha_1)] + (1-\alpha_1) \\ &= \alpha_1^2 p_0 + (1-\alpha_1^2) \end{aligned}$$

If this procedure is continued k times the result obtained would be

(3.4.3) $$p_k = Q_1^k p = \alpha_1^k p_0 + (1-\alpha_1^k) \quad .$$

This general form may be proven to be correct by using mathematical induction.

The third assumption of all the words being equally difficult can be satisfied very easily when the words are nonsense syllables. Both Glaze (1928) and Kruger (1934) have computed meaningfulness of nonsense syllables. By picking out syllables that are equally meaningful the syllables would be approximately equal in difficulty.

When using monosyllables the difficulty of a word would depend on each subject's background. There is no criterion that can be used to rate monosyllables on difficulty. This does not mean the model cannot be used, but if the model does not fit the data very well the experimenter should be aware this assumption may have been incorrect. Also it should be noted that this assumption implies that primacy and recency have no effect.

The fifth assumption means that the subjects are given as many trials as they need in order to learn the list.

3.5 Estimation of the Parameter p_0.

The initial trial is equivalent to N binomial trials with a probability p_0 of a success, where N is the total number of words to be recalled. Let $x_{i,0} = 1$ if the i^{th} word is recalled on the initial trial, and $x_{i,0} = 0$ if it is not. Using Fryer (1966),

$$(3.5.1) \qquad \hat{p}_0 = \frac{1}{N} \sum_{i=1}^{N} x_{i,0}$$

would be an unbiased maximum likelihood estimator of p_0 with variance

$$(3.5.2) \qquad \sigma^2(\hat{p}_0) = \frac{\hat{p}_o(1-\hat{p}_o)}{N} \quad .$$

For a quick and easy way to obtain an estimate of p_0 the above estimates can be used.

A better estimate of p_0 can be made by using more information. Since the non-recall of a word doesn't change its probability of being recalled on the trial, the data for each word can be used to estimate p_0. For each word the number of trials preceded

entirely by zero recalls can be obtained from the data. Let N_0 be the total number of word trials which are preceded entirely by zero recalls. Using Mood and Graybill (1963) the probability of obtaining a value x of N_0 can be found from the negative binomial distribution, and is given by

$$p(N_o=x) = f(x) = \binom{x-1}{N-1} (1-p_o)^{x-N} p_o^{N} .$$

To maximize the likelihood function $f(N_0)$ the logarithm can be differentiated with respect to p_o, and set equal to zero.

$$L^* = \log L = \log \binom{N_o-1}{N-1} + (N_0-N) \log (1-p_o) + N \log p_o$$

$$\frac{\partial L^*}{\partial p_o} = - \frac{N_o-N}{1-p_o} + \frac{N}{p_o} = 0$$

From which the maximum likelihood estimate of p_o is obtained as

(3.5.3) $\hat{p}_o = N/N_0$.

This estimate is not unbiased, but Girshick, Mosteller, and Savage (1946) have shown that when N is fixed and N_0 is varied the estimator

$$\hat{p}_0 = \frac{N-1}{N_o-1}$$

is unbiased. For large N, however (3.5.3) can be used. Bush and Mosteller (1955) showed the asymptotic variance of (3.5.3) to be

(3.5.4) $$\sigma^2(\hat{p}_o) = \frac{\hat{p}_o^2(1-\hat{p}_o)}{N} .$$

This variance is smaller than the variance of (3.5.2) when p_o is less than one, because of the extra p_o term in (3.5.4).

3.6 Estimate of the Parameter α_1.

After n trials there will be 2^n possible and different sequences of recalls and non-recalls of a word. Let $\overline{q}_{k,n}$ be the probability of non-recall of a word in the kth sequence on the nth trial. In the same way (3.4.1) and (3.4.2) were derived, the probabilities for recall and non-recall of a word of the kth sequence on trial n+1 are given by

$$Q_1\overline{p}_{k,n}=\alpha_1\overline{p}_{k,n}+(1-\alpha_1)$$

and $$Q_2\overline{p}_{k,n}=\overline{p}_{k,n} \quad .$$

Using the fact $Q_i\overline{p}_{k,n}=1-\tilde{Q}_i\overline{q}_{k,n}$, then the above equations can be rewritten as

$$\tilde{Q}_1\overline{q}_{k,n}=\alpha_1\overline{q}_{k,n}$$

and $$\tilde{Q}_2\overline{q}_{k,n}=\overline{q}_{k,n} \quad .$$

A word in the kth sequence is recalled with probability $1-\overline{q}_{k,n}$, and if the word is recalled on trial n it has probability $\alpha_1\overline{q}_{k,n}$ of not being recalled on trial n+1. A word in the kth sequence is not recalled with probability $\overline{q}_{k,n}$, and if the word is not recalled on trial n it has probability $\overline{q}_{k,n}$ of not being recalled on trial n+1. The mean value of $\overline{q}_{k,n+1}$ by definition is

$$E(\overline{q}_{k,n+1}) = \alpha_1\overline{q}_{k,n}(1-\overline{q}_{k,n}) + \overline{q}_{k,n}\ \overline{q}_{k,n}.$$

To find the mean value over the entire population of words, denoted by $V_{1,n+1}$, all the values $\overline{q}_{k,n}$ are summed, each weighted by its probability of occurence $Q_{k,n}$. So

$$V_{1,n+1} = \sum_{k=1}^{2^n} (\alpha_1 \bar{q}_{k,n}(1-\bar{q}_{k,n}) + \bar{q}^2_{k,n}) \cdot Q_{k,n}$$

$$= \alpha_1 \sum_{k=1}^{2^n} \bar{q}_{k,n} Q_{k,n} + (1-\alpha_1) \sum_{i}^{2^n} \bar{q}^2_{k,n} Q_{k,n}$$

$$(3.6.1) \qquad = \alpha_1 V_{1,n} + (1-\alpha_1) V_{2,n}$$

where $V_{2,n}$ is the second moment of the $\bar{q}_{k,n}$ values about the origin. Equation (3.6.1) does not give an exact solution, because of the $V_{2,n}$ term. Bush and Mosteller (1955) using several approximations found that

$$(3.6.2) \qquad \bar{T}_2 = E(T_2) \approx -\frac{1}{1-\alpha_1} \ln(1-q_o) \quad .$$

where $\bar{T}_2$ is the mean total number of non-recalls. Using (3.6.2) gives

$$(3.6.3) \qquad \alpha_1 = 1 - \frac{-\ln \hat{p}_o}{\hat{\bar{T}}_2}$$

By counting the number of non-recalls actually made in the experiment the quantity $\bar{T}_2$ can be estimated. The value of p_o can be estimated by using either method described in Section 3.5. Knowing these estimates the value of α_1 can be estimated quite easily by (3.6.3).

By knowing only the estimated values of p_o and α_1 the data of a free recall experiment can easily be summarized by the Bush and Mosteller model.

Bush and Mosteller (1955) have given other ways to estimate α_1 for special cases, i.e. when q_o equals one. It is not

worthwhile in this paper to consider the methods, because (3.6.3) can be used for the special cases, and the amount of calculation to apply the methods is greater than when (3.6.3) is used.

4. The Miller and McGill Stochastic Model.

Miller and McGill (1952) developed a stochastic model that is closely related to the linear model. Using the same assumptions that were used in the linear model other quantities of interest can be studied by using the Miller and McGill model.

4.1 Definitions and Terms of the Model.

Miller and McGill (1952) classified words according to which state they were in, where a word which had been recalled exactly k times on the preceding trials was said to be in state A_k. The probability that a word was in state A_k on trial n was denoted by $p(A_k,n)$.

4.2 The Difference Equation and Its Solution.

A word can get into state A_k on trial n+1 in only two ways. Either a word is in state A_k on trial n and it is not recalled on trial n+1, or the word is in state A_{k-1} on trial n and it is recalled. So the difference equation to represent this is

(4.2.1) $\quad p(A_k,n+1) = p(A_k,n)(1-p_k) + p(A_{k-1},n)p_{k-1}$

where p_k is the probability that a word will be recalled after k recalls, and is given by (3.4.3).

To obtain the general solution of (4.2.1), the system of equations of (4.2.1) is written in matrix form

$$(4.2.2)\quad \begin{bmatrix} 1-p_o & 0 & 0 & 0 & \cdots \\ p_o & 1-p_1 & 0 & 0 & \cdots \\ 0 & p_1 & 1-p_2 & 0 & \cdots \\ 0 & 0 & p_2 & 1-p_3 & \cdots \\ \vdots & \vdots & \vdots & & \vdots\vdots\vdots \end{bmatrix} \cdot \begin{bmatrix} p(A_0,n) \\ p(A_1,n) \\ p(A_2,n) \\ p(A_3,n) \\ \vdots \end{bmatrix} = \begin{bmatrix} p(A_0,n+1) \\ p(A_1,n+1) \\ p(A_2,n+1) \\ p(A_3,n+1) \\ \vdots \end{bmatrix} .$$

Let T be the first matrix in (4.2.2), the matrix of transition probabilities. Let d_n and d_{n+1} be the column vectors of the state probabilities on trial n and trial n+1 respectively, also in (4.2.2). So (4.2.2) may be written as

$$Td_n = d_{n+1}$$

The state probabilities on trial one are given by $Td_o = d_1$. The state probabilities on trial two are given by $Td_1 = d_2$, or

$$Td_1 = T(Td_o) = T^2 d_o = d_2 \quad .$$

Continuing this procedure, it is apparent that the state probabilities on trial n are given by

$$T^n d_o = d_n \quad .$$

By Rao (1965), the semi-matrix T can be written as

(4.2.3) $\quad T = S\ D\ S^{-1}$

where D is an infinite diagonal matrix with the same elements on its diagonal as are on the main diagonal of T, and with the remaining elements being zero. So T^2 may be written as

$$T^2 = (SDS^{-1})(SDS^{-1}) = SD^2S^{-1} \quad .$$

In general, then

$$T^n = SD^nS^{-1} \quad .$$

Since D is a diagonal matrix, D^n is obtained by taking the nth power of every diagonal element of D. Rewrite (4.2.3) as

$$TS = SD \quad .$$

The diagonal elements of S are arbitrary, so let $S_{ii} = 1$.

$$\begin{bmatrix} 1-p_o & 0 & 0 & \cdots \\ p_o & 1-p_1 & 0 & \cdots \\ 0 & p_1 & 1-p_2 & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & \cdots \\ S_{21} & 1 & 0 & \cdots \\ S_{31} & S_{32} & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \cdots \\ S_{21} & 1 & 0 & \cdots \\ S_{31} & S_{32} & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} 1-p_o & 0 & 0 & \cdots \\ 0 & 1-p_1 & 0 & \cdots \\ 0 & 0 & 1-p_2 & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

The S_{ij} terms can be solved for term by term. The matrix S turns out to be equal to

$$\begin{bmatrix} 1 & 0 & 0 & \cdots \\ \dfrac{p_0}{p_1-p_o} & 1 & 0 & \cdots \\ \dfrac{p_op_1}{(p_1-p_o)(p_2-p_o)} & \dfrac{p_1}{p_2-p_1} & 1 & \cdots \\ \dfrac{p_op_1p_2}{(p_1-p_o)(p_2-p_o)(p_3-p_o)} & \dfrac{p_1p_2}{(p_2-p_1)(p_3-p_1)} & \dfrac{p_2}{p_3-p_2} & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} .$$

Taking the inverse of S gives S^{-1} which turns out to be

$$\begin{bmatrix} 1 & 0 & 0 & 0\cdots \\ \frac{p_o}{p_o-p_1} & 1 & 0 & 0\cdots \\ \frac{p_op_1}{(p_o-p_2)(p_1-p_2)} & \frac{p_1}{p_1-p_2} & 1 & 0\cdots \\ \frac{p_op_1p_2}{(p_o-p_3)(p_1-p_3)(p_2-p_3)} & \frac{p_1p_2}{(p_1-p_3)(p_2-p_3)} & \frac{p_2}{p_2-p_3} & 1\cdots \\ \vdots & \vdots & \vdots & \vdots\vdots \end{bmatrix}.$$

So the first column of T^n turns out to be

$$\begin{bmatrix} (1-p_o)^n \\ p_o\left|\frac{(1-p_o)^n}{(p_1-p_o)} + \frac{(1-p_1)^n}{p_o-p_1}\right| \\ p_op_1\left|\frac{(1-p_o)^n}{(p_1-p_o)(p_2-p_o)} + \frac{(1-p_1)^n}{(p_o-p_1)(p_2-p_1)} + \frac{(1-p_2)^n}{(p_o-p_2)(p_1-p_2)}\right| \\ \vdots \end{bmatrix}.$$

The reason why only the first column of T^n was found is because d_o is just the column vector $(1,0,0,\ldots)$. So, T^nd_o involves only the first column of T^n and thus the general solution of (4.2.1) is

(4.2.3) $$p(A_o,n) = (1-p_o)^n \quad , \quad \text{for } k = 0$$

$$p(A_k,n) = p_o p_1 \cdots p_{k-1} \sum_{i=1}^{k} \frac{(1-p_i)^n}{\prod_{\substack{j=0 \\ j\neq i}}^{k} (p_j - p_i)} \quad , \quad k > 0$$

Using the results of the Bush and Mosteller linear model the parameters p_o and α_1 can be estimated. The set of p_k's can be found by (3.4.3). Substituting for the p_k's in (4.2.3) the following is obtained.

(4.2.4)
$$p(A_o,n) = (1-p_o)^n$$
$$P(A_k,n) = (1-p_o)^n \prod_{i=0}^{k-1} \frac{(1-(1-p_o)\alpha_1^i)(1-\alpha_1^{n-i})}{1-\alpha_1^{i+1}}$$

4.3 Expected Number of Times a Word is Recalled.

Let E(k,n) be the expected number of times a word is recalled up to and including trial n. By definition E(k,n) is

(4.3.1)
$$E(k,n) = \sum_{k=0}^{n} kp(A_k,n) \quad .$$

Let r_{n+1} be the expected number of words recalled on trial n+1. By definition r_{n+1} is

(4.3.2)
$$r_{n+1} = E(k,n+1) - E(k,n) \quad .$$

or
$$r_{n+1} = \sum_{k=0}^{n+1} kp(A_k,n+1) - \sum_{k=0}^{n} kp(A_k,n) \quad .$$

Using (4.2.2) the first summation is rewritten so that

$$r_{n+1} = \sum_{k=0}^{n} kp(A_k,n)(1-p_k) + \sum_{k=1}^{n+1} kp(A_{k-1},n)p_{k-1} - \sum_{k=0}^{n} kp(A_k,n) \quad .$$

or
$$r_{n+1} = \sum_{k=0}^{n} kp(A_k,n) - \sum_{k=0}^{n} kp(A_k,n)p_k + \sum_{k=0}^{n} (k+1)p(A_k,n)p_k$$
$$- \sum_{k=0}^{n} kp(A_k,n) \quad .$$

So, $r_{n+1} = \sum_{k=0}^{n} (k+1)p(A_k,n)p_k - \sum_{k=0}^{n} kp(A_k,n)$

or $r_{n+1} = \sum_{k=0}^{n} p_k p(A_k,n)$.

The Bush and Mosteller model is used more to summarize the data, and not for prediction. Miller and McGill's model can be used for prediction. By comparing the predictions made and the data it can be seen how well the model works.

5. The Waugh and Smith Stochastic Model

Waugh and Smith's (1962) model uses a Markovian process with an absorbing state. For a general discussion of Markovian models in psychology see Miller (1952), Kao (1953), and Goodman (1953).

5.1 Definitions and Terms.

Waugh and Smith (1962) defined three processes that were named labeling, selecting, and fixing. The process of labeling was equivalent to a word acquiring a mnemonic tag. For a word to be recalled it must be labeled, but if a word is labeled it doesn't mean the word will be recalled. Labeling occurs with probability λ on any trial, and is irreversible. In other words, once a word is labeled it stays labeled. The second process of selecting is equivalent to rehearsing a word. Selecting a word is assumed to occur with probability σ on each trial. For a word to be recalled for the first time on a given trial the word must have been labeled on that trial or on some previous trial, and it must be selected on that trial. The third process, fixing a word, is assumed to occur with probability ϕ on any trial in

which a word is recalled. Once a word is fixed it is recalled on every subsequent trial. If it is not fixed the word is forgotten, and the word must be selected again with probability σ.

5.2 States of the Waugh and Smith Model.

A word may be in any one of five states after a given trial. In state one the word hasn't been labeled yet. A word in state two has been labeled, but not selected yet. For state three the word has been labeled and selected, but not as yet fixed. In state three the word was recalled, because it had been labeled and selected. In state four the word has been recalled but not fixed on some previous trial, and it was not selected on the given trial. A word in state five has been fixed. State five is an absorbing state. The trials are continued until perfect retention is obtained.

Let $P_{n,j}$ be the probability of a word being in state j on trial n. By considering how a word can get to one state from other states the following equations may be written.

$$P_{n,1} = (1-\lambda)P_{n-1,1}$$

$$P_{n,2} = (1-\sigma)P_{n-1,2}+\lambda(1-\sigma)P_{n-1,1}$$

$$P_{n,3} = \sigma(1-\phi)(P_{n-1,2}+P_{n-1,3}+P_{n-1,4})+\sigma\lambda(1-\phi)P_{n-1,1}$$

$$P_{n,4} = (1-\sigma)(P_{n-1,3}+P_{n-1,4})$$

$$P_{n,5} = P_{n-1,5}+\sigma\phi(P_{n-1,2}+P_{n-1,3}+P_{n-1,4})+\lambda\sigma\phi P_{n-1,1}$$

The system of equations may be written in matrix notation.

$$\begin{bmatrix} (1-\lambda) & 0 & 0 & 0 & 0 \\ \lambda(1-\sigma) & 1-\sigma & 0 & 0 & 0 \\ \sigma\lambda(1-\phi) & \sigma(1-\phi) & \sigma(1-\phi) & \sigma(1-\phi) & 0 \\ 0 & 0 & 1-\sigma & 1-\sigma & 0 \\ \sigma\phi\lambda & \sigma\phi & \sigma\phi & \sigma\phi & 1 \end{bmatrix} \begin{bmatrix} P_{n-1,1} \\ P_{n-1,2} \\ P_{n-1,3} \\ P_{n-1,4} \\ P_{n-1,5} \end{bmatrix} = \begin{bmatrix} P_{n,1} \\ P_{n,2} \\ P_{n,3} \\ P_{n,4} \\ P_{n,5} \end{bmatrix}$$

Let, as before in Section (4.2), T be the transpose of the matrix of transition probabilities, and d_n denote the column vector of state probabilities on trial n. Using the same method as in (4.2) then

$$T^n d_o = d_n \quad .$$

The initial vector d_o' is the vector (1,0,0,0,0), because all of the words start in state one. Therefore, only the first column of the matrix T^n needs to be found to find the elements of the vector d_n. If T is multiplied by itself a few times a pattern soon develops. The elements of the first column of T^n can be written by comparing terms. Thus, the elements of d_n turn out to be

$$P_{n,1} = (1-\lambda)^n$$

$$P_{n,2} = \frac{\lambda(1-\sigma)}{\sigma-\lambda} \left((1-\lambda)^n - (1-\sigma)^n \right)$$

$$\text{(5.2.1)} \quad P_{n,3} = \frac{\lambda\sigma(1-\phi)}{(\lambda-\sigma\phi)} \left((1-\sigma\phi)^n - (1-\lambda)^n \right)$$

$$P_{n,4} = \frac{\lambda(1-\sigma)}{\sigma-\phi} (1-\sigma\phi)^n + \frac{\lambda}{\sigma-\lambda} (1-\sigma)^{n+1} - \frac{\sigma(1-\sigma)\lambda(1-\phi)}{(\sigma-\lambda)(\lambda-\sigma\phi)} (1-\lambda)^n$$

$$P_{n,5} = 1-(1-\sigma\phi)^{n+1} - \frac{\sigma\phi}{\lambda-\sigma\phi}\left((1-\sigma\phi)^{n+1} - (1-\lambda)^{n+1} \right) \quad .$$

Once a word has been recalled for the first time it can never return to state one or state two. The probability $R_{j,k}$ of a word being in state k after j trials from the first recall trial is given by the matrix equation

$$(5.2.2) \quad \begin{bmatrix} \sigma(1-\phi) & \sigma(1-\phi) & 0 \\ 1-\sigma & 1-\sigma & 0 \\ \sigma\phi & \sigma\phi & 1 \end{bmatrix} \begin{bmatrix} R_{j-1,3} \\ R_{j-1,4} \\ R_{j-1,5} \end{bmatrix} = \begin{bmatrix} R_{j,3} \\ R_{j,4} \\ R_{j,5} \end{bmatrix}$$

Let Q be the first matrix in (5.2.2) the transpose of the matrix of transition probabilities. Let S_j and S_{j+1} be the column vectors of the state probabilities on trial j and j+1 respectively, also in (5.2.2). Then using the same procedure as in Section (4.2)

$$Q^j S_o = S_j \quad .$$

The S_o is equal to $(1-\phi,\ 0,\ \phi)'$, because a proportion ϕ of the words are fixed on the trial on which they are first recalled, while a proportion $1-\phi$ are selected but not fixed. Those selected but not fixed go into state three. If Q is multiplied by itself a few times a pattern soon develops. Using this pattern the elements of Q^j can be easily found.

$$Q^j S_o = \begin{bmatrix} \sigma(1-\phi)(1-\sigma\phi)^{j-1} & \sigma(1-\phi)(1-\sigma\phi)^{j-1} & 0 \\ (1-\sigma)(1-\sigma\phi)^{j-1} & (1-\sigma)(1-\sigma\phi)^{j-1} & 0 \\ 1-(1-\sigma\phi)^j & 1-(1-\sigma\phi)^j & 1 \end{bmatrix} \begin{bmatrix} 1-\phi \\ 0 \\ \phi \end{bmatrix} = \begin{bmatrix} R_{j,3} \\ R_{j,4} \\ R_{j,5} \end{bmatrix} = S_j$$

Therefore, the solution for $R_{j,k}$ is

$$R_{j,3} = \sigma(1-\phi)^2(1-\sigma\phi)^{j-1}$$

(5.2.3) $$R_{j,4} = (1-\sigma)(1-\phi)(1-\sigma\phi)^{j-1}$$

$$R_{j,5} = 1 - (1-\phi)(1-\sigma\phi)^{j} \quad .$$

The probability R_j that a word will be recalled after j trials from the first recall is

$$R_j = 1 - R_{j,4} \quad .$$

Using (5.2.3), then R_j is

(5.2.4) $$R_j = 1 - (1-\sigma)(1-\phi)(1-\sigma\phi)^{j-1} \quad .$$

5.3 Estimation of the Parameters.

Let the probability of first recall on trial n be F_n. This probability is found by considering Pr(1st recall by nth trial) = Pr(1st recall on n^{th} trial or 1st recall by the (n-1)th trial). This statement may be rewritten as Pr(1st recall by nth trial) = Pr(1st recall on trial n) + Pr(1st recall by the (n-1)th trial), or Pr(1st recall on trial n) = Pr(1st recall by nth trial) - Pr(1st recall by the (n-1)th trial) = Pr(not yet recalled by the (n-1)th trial) - Pr(not yet recalled by nth trial). Thus, $F_n = P_{n-1,2} + P_{n-1,1} - P_{n,2} - P_{n,1}$ or using (5.2.1) F_n is equal to

(5.3.1) $$F_n = \frac{\sigma\lambda}{\sigma-\lambda}\left((1-\lambda)^n - (1-\sigma)^n \right) \quad .$$

The F_n can be estimated from the data for various values of n. Let $x_{i,n} = 0$ if the ith word is not recalled on trial n, or if it has been recalled before the given trial. Let $x_{1,n} = 1$ if the

first recall of the word occurs on trial n. Then F_n would be estimated by

$$\hat{F}_n = \frac{1}{N} \sum_{i=1}^{N} x_{i,n} \quad .$$

The parameters σ and λ are estimated by finding the minimum chi-square estimates of σ and λ that best fit the data of (5.3.1).

Let R_n be the probability of recall on trial n. When a word is recalled on a trial it must be in either state three or state five. Therefore, R_n is given by

$$R_n = P_{n,5} + P_{n,3} \quad .$$

If (5.2.1) is used, then R_n can be written as

$$(5.3.2) \qquad R_n = 1 - (1-\sigma\phi)^n - \frac{\sigma(\phi-\lambda)}{\lambda-\sigma\phi} \left((1-\sigma\phi)^n - (1-\lambda)^n \right) \quad .$$

The quantity R_n can be estimated from the data by

$$\hat{R}_n = \frac{1}{N} \sum_{i=1}^{N} y_{i,n} \quad ,$$

where $y_{i,n} = 1$ if the ith word is recalled on trial n, and $y_{i,n} = 0$ if it is not recalled. Using the estimated values of σ, λ, and R_n the least-squares estimate of ϕ is found. This is the estimate of ϕ that is used in the model.

6. The Cowan Stochastic Model

The Cowan model is unlike the previous models discussed, because this model considers the effect of associative connections

between words in recalling them. Bousfield (1953) showed that there is a tendency for some words to cluster together when the list is recalled. When a certain type of word is recalled the remaining words that have a high association value to the word recalled are likely to have a higher probability of recall than words of lower association strength. The entire list is given a number of times, each time in a different order. After a number of presentations the subject is asked to recall as many words as he can. The Cowan model predicts the kinds of words that will appear in a given recall position.

6.1 Definitions and Terms.

Cowan (1966) considers a list of stimulus words that could be divided into two groups. One group is denoted as Category C_1 and the other as Category C_2. An example would be if C_1 consisted of tree names, and C_2 consisted of words that were selected randomly. The strength of C_1 or C_2 is defined in terms of the associative connections which exist between its members.

There are four sets of associative interconnections. There are two within-category associations, $(C_1 \rightarrow C_1)$ and $(C_2 \rightarrow C_2)$. There are also the between-category associations, $(C_1 \rightarrow C_2)$ and $(C_2 \rightarrow C_1)$. If the first word recalled is a C_2 word, then the probability of recalling a C_1 word next would be

$$P(C_1 | C_2) = \frac{M(C_2 \rightarrow C_1)}{M(C_2 \rightarrow C_1) + M(C_2 \rightarrow C_2)}$$

where $M(\cdot)$ is a measure of association of $(\cdot)$. The probabilities $P(C_2|C_1)$, $P(C_1|C_1)$ and $P(C_2|C_2)$ would be defined similarly.

6.2 Estimation of Association Strengths.

A method suggested by Pollio (1963) to measure the within-category and the between-category association strengths may be used. The method used is to set up matrices of C_1xC_1, C_2xC_2, and C_1xC_2. In each cell the association strength between the corresponding words is entered. Let c_1 and c_2 be the total number of words in C_1 and C_2 respectively. Let $C_1(i)$ be the ith word in C_1. The association strengths for selected word lists can be found in Palemo and Jenkins (1964).

$$
\begin{array}{c|c|c|c|c|}
 & C_1(1) & C_1(2) & \cdots & C_1(c_1) \\
\hline
C_1(1) & 0 & & & \\
\hline
C_1(2) & & 0 & & \\
\hline
\vdots & & & \ddots & \\
\hline
C_1(c_1) & & & & 0 \\
\hline
\end{array}
\quad \Sigma\Sigma = \alpha_1
$$

The sum of the entries of the C_1xC_1, C_2xC_2, and C_1xC_2 matrices are symbolized by α_1, α_2, and β respectively. The mean association value between any C_1 word occurring first in recall and the remaining C_1 words is given by α_1/c_1, where c_1 represents the total number of words in the C_1 category. Similarly the mean associative value estimate of a C_2 word leading to another C_2 would be α_2/c_2. For a C_1 word leading to a C_2 word the estimate would be β/c_1, and for a C_2 word leading to a C_1 would be β/c_2.

In this model the words within a category are assumed to be indistinguishable from each other. Thus, the association between any pair in a category is the same as any other pair and the mean association strength is used to estimate this value.

The probability of a C_1 word on first recall on a given trial followed by another C_1 word would be

$$(6.2.1)\quad P(C_1|C_1) = \frac{M(C_1 \to C_1)}{M(C_1 \to C_1)+M(C_1 \to C_2)} = \frac{\alpha_1/c_1}{\alpha_1/c_1+\beta/c_1} = \frac{\alpha_1}{\alpha_1+\beta} \quad .$$

The probabilities $P(C_1|C_2)$, $P(C_2|C_2)$, and $P(C_2|C_1)$ would be defined similarly.

6.3 The Non-Markovian Process of the Cowan Model.

The probabilities of (6.2.1) change on the next recall of the same trial, because once a word has been recalled it will not be a possible response for the next word recalled. Thus, the within-category mean association value and the between-category mean association value are reduced. Since it was assumed that all words are indistinguishable in a category, the association strength between each pair is equal in value to the association strength between every other pair.

The mean association strength between each pair in C_1 is given by

$$(6.3.1)\qquad \alpha_1/c_1(c_1-1) \quad .$$

So, the new within-category mean association value encountered by the second C_1 word would be

$$\frac{\alpha_1}{c_1} - \frac{\alpha_1}{c_1(c_1-1)} \quad \text{or} \quad \frac{\alpha_1}{c_1^2-c_1}\,(c_1-2) \quad .$$

Each time another word from C_1 is recalled the mean within-association value is reduced by the amount given by (6.3.1). When a total of r_1 words from C_1 have been recalled the mean association value encountered by the next C_1 word is given by

$$(6.3.2) \qquad M(C_1 \to C_1) = \frac{\alpha_1}{c_1^2 - c_1}(c_1 - r_1 - 1) \quad .$$

Similarly, when a total of r_2 words from C_2 have been recalled the mean within-association value left for the next C_2 word is given by

$$(6.3.3) \qquad M(C_2 \to C_2) = \frac{\alpha_2}{c_2^2 - c_2}(c_2 - r_2 - 1) \quad .$$

The mean between-association strength for a C_1 word and a C_2 word would be given by

$$(6.3.4) \qquad \frac{\beta}{c_1 c_2}$$

Each time a C_2 word is recalled the mean association strength is reduced by the amount given in (6.3.4). So, after r_2 words are recalled from C_2 the remaining mean association strength left for the next C_1 word is given by

$$(6.3.5) \qquad M(C_1 \to C_2) = \frac{\beta}{c_1 c_2}(c_2 - r_2) \quad .$$

Similarly for r_1 words recalled from C_1 the mean association strength of a C_2 leading to a C_1 word would be equal to

(6.3.6) $$M(C_2 \to C_1) = \frac{\beta}{c_1 c_2}(c_1 - r_1) \quad .$$

Thus, using the above equations for M(.) and the equations for P(.), then

a. $$P(C_1|C_1) = \frac{\alpha_1 c_2(c_1-r_1-1)}{\alpha_1 c_2(c_1-r_1-1)+\beta(c_1-1)(c_2-r_2)}$$

b. $$P(C_2|C_1) = \frac{\beta(c_1-1)(c_2-r_2)}{\alpha_1 c_2(c_1-r_1-1)+\beta(c_2-1)(c_2-r_2)}$$

(6.3.7)

c. $$P(C_2|C_2) = \frac{\alpha_2 c_1(c_2-r_2-1)}{\alpha_2 c_1(c_2-r_2-1)+\beta(c_2-1)(c_1-r_1)}$$

d. $$P(C_1|C_2) = \frac{\beta(c_2-1)(c_1-r_1)}{\alpha_2 c_1(c_2-r_2-1)+\beta(c_2-1)(c_1-r_1)}$$

The process can be in two states, C_1 or C_2. The transitional probabilities are functions of the number of each type of word recalled, and so this is a Non-Markovian process.

6.4 The Transition Matrix.

By redefining the states to represent the type of word and the number of words of each type recalled, by Feller (1957), the process can be treated as a Markov chain. Let $C_i(m,n)$ be the state in which a C_i word has just been given with m C_1 words and n C_2 words previously recalled.

The probabilities for the transition matrix are found from (6.3.7). For example, consider the probability of going from state $C_1(i,n)$ to state $C_1(i+1,n)$. Using (6.3.7)a. with $r_1 = i$ the $P(C_1(i,n)|C_1(i+1,n))$ is calculated and substituted into the transition matrix.

The transition matrix P can be arranged into sets, such that when one set is reached the process cannot enter the states located below it, and after a set has been entered it is immediately left. All sets then are transient except the set representing complete recall, and those states would be absorbing. An example of how matrix P would be arranged is given in Table 1. These sets contain all the possible states involved in recall of lengths denoted by the set number. The states are numbered to conserve space and divided into sets labeled I, II, III, etc. Matrices of the form in Table 1 are submatrices of the matrix P. Let Q be any submatrix formed this way. The sets in matrix Q are transient. Kemeny and Snell (1960) have proved a matrix H which gives the probabilities that a process will ever go from any transient state to any other transient state is given by

$$H = (N-I)\, N_{dg}^{-1}$$

where $N = (I-Q)^{-1}$, and N_{dg} is a diagonal matrix whose elements are the same as the diagonal elements of N. The matrix Q has only non-zero elements below the diagonal. So the matrix N would have ones on the diagonal. Thus N_{dg}^{-1} would be the identity matrix, so

$$H = (I-Q)^{-1} - I. \tag{6.4.1}$$

For example, the probability of starting in state $C_1(0,0)$ and ending in state $C_2(3,2)$ in the sixth recall position can be found in the matrix H. The matrix Q would include the sets I through VI.

TABLE 1

TRANSITION MATRIX P FOR THE FIRST FOUR WORDS RECALLED

(X's signify nonzero entries)

States	State No.	State No.																			
		IV								III						II				I	
		1	2	3	4	5	6	7	8	1	2	3	4	5	6	1	2	3	4	1	2
	IV																				
$C_1(3,0)$	1																				
$C_2(3,0)$	2																				
$C_1(2,1)$	3																				
$C_2(2,1)$	4																				
$C_1(1,2)$	5																				
$C_2(1,2)$	6																				
$C_1(0,3)$	7																				
$C_2(0,3)$	8																				
	III																				
$C_1(2,0)$	1	X	X																		
$C_2(2,0)$	2			X	X																
$C_1(1,1)$	3			X	X																
$C_2(1,1)$	4					X	X														
$C_1(0,2)$	5					X	X														
$C_2(0,2)$	6							X	X												
	II																				
$C_1(1,0)$	1									X	X										
$C_2(1,0)$	2											X	X								
$C_1(0,1)$	3											X	X								
$C_2(0,1)$	4													X	X						
	I																				
$C_1(0,0)$	1															X	X				
$C_2(0,0)$	2																	X	X		

6.5 Adjustments Made for the Model.

When the set of data and the model were compared it was found necessary to make some adjustments. It was found by Cowan that a better fit was obtained by using c_1 and c_2 as the mean number of C_1 and C_2 words recalled respectively. Cowan thought the reason for having to redefine c_1 and c_2 was because the subject received, organized, and recalled completely only a limited number of items on the list. Finally, α_2 was made free and a family of curves were generated. The value of α_2 was picked which gave the best fit to the data. Cofer and Reicher (1964), and Puff (1964) demonstrated that when words in a category appear together in the list presented, they will tend to appear together in recall. Thus, the occurrence of items together in the list might increase the association between them, and this would increase the value of α.

7. Summary

Using (3.5.3) and (3.6.3) the values of p_o and α_1 can be estimated. Knowing only these two values the data of a free recall experiment can be summarized by Bush and Mosteller's model. Knowing the estimates of α_1 and p_o the probability p_k of recalling a word after k recalls can be found by using (3.4.3) of the linear model.

Using Miller and McGill's model and the values of p_o and α_1 estimated by the Bush and Mosteller model an experimenter can

find the probability of a word being recalled exactly k times on trial n by using (4.2.4). Using Miller and McGill's model the expected number of times a word is recalled up to and including trial n can be found by using (4.3.1). Another quantity of interest is the expected number of words recalled on trial n+1, which can be found by using (4.3.2).

If the process of labeling, selecting, and fixing of a word are considered, the Waugh and Smith model may be used. With their model the probability R_j that a word will be recalled after j trials from the first recall can be found by using (5.2.4). By estimating the probability F_n of first recall on the nth trial the values of σ and λ can be found by using the best minimum chi-square fit to (5.3.1). By estimating the probability R_n of recalling a word on trial n from the data and using the estimates of σ and λ, the least-squares estimate of ϕ is found using (5.3.2). If the values of σ, λ, and ϕ are already known, say from a previous and similar experiment, the probability of first recall on trial n and the probability of recall on trial n can be found by (5.3.1) and (5.3.2) respectively.

Cowan's model is used when an experimenter wishes to consider the effect of associations between words. The model is limited to the case where the words in a list can be put into two categories. Once a measure of association is found between categories or within categories various probabilities can be found. Using (6.3.7) the probability of a word from a category following

a word from the same category or the other category can be found given that r_1 words have been recalled from C_1 and r_2 words from C_2. A transition matrix can be formed by letting the states represent the type of word and the number of words of each type recalled. Using (6.4.1) the probability of starting in a state and ending in a certain state can be found.

By comparing the predictions of the models and the data obtain the experimenter can determine which model best fits his experiment. With the parameter values known the data can be summarized. Individual subjects can be compared easily, and the effects of changing the number of words in the list or speed of presentation of words can be measured readily in terms of the parameters.

8. Acknowledgements

This writer would like to thank Dr. W. J. Conover for all the help which he has given concerning this paper, and also to thank the Psychology Department for letting me use their library.

REFERENCES

Atkinson, R. C., Bower, G. H. and Crothers, E. J. (1965). An Introduction to Mathematical Learning Theory. Wiley, New York.

Bush, R. R. and Mosteller, F. (1951). A mathematical model for simple learning. Psychol. Rev. **85** 313-323.

Bush, R. R. and Mosteller, F. (1953). A stochastic model with applications to learning. Ann. Math. Statist. **24** 559-584.

Bush, R. R. and Mosteller, F. (1955). Stochastic Models for Learning. Wiley, New York.

Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. J. Gen. Psychol. **49** 299-240.

Cofer, C. N. and Reicher, G. M. (1964). Clustering in free recall as a function of certain methodological variations. Technical Report No. 3. Office of Naval Research and Pennsylvania State University.

Cowan, T. M. (1966). A Markov model for order of emissions in free recall. J. of Math. Psych. **3** 470-483.

Estes, W. K. (1950). Toward a statistical theory of learning. Psychol. Rev. **57** 97-107.

Feller, W. (1957). An Introduction to Probability Theory and Its Application. Wiley, New York.

Fryer, H. C. (1966). Concepts and Methods of Experimental Statistics. Allyn and Bacon, Boston.

Girshick, M. A., Mosteller, F. and Savage, L. J. (1946). Unbiased estimates for certain binomial sampling problems with applications. Ann. Math. Statist. **17** 13-23.

Glaze, J. A. (1928). The association value of nonsense syllables. J. Genet. Psychol. **35** 255-269.

Goodman, L. A. (1953). A further note on "Finite Markov processes in psychology". Psychometrika. **18** 245-247.

Kao, R. C. W. (1953). Note on Miller's "Finite Markov processes in psychology". Psychometrika. **18** 241-243.

Kemeny, J. G. and Snell, J. L. (1960). Finite Markov Chains. Van Nostrand, Princeton.

Kruger, W. C. F. (1934). The relative difficulty of nonsense syllables. J. Exp. Psychol. **17** 145-153.

Miller, G. A. (1952). Finite Markov processes in psychology. Psychometrika. **17** 149-168.

Miller, G. A. (1964). Mathematics and Psychology. Wiley, New York.

Miller, G. A. and McGill, W. J. (1952). A statistical description of verbal learning. Psychometrika. **17** 369-396.

Mood, A. M. and Graybill, F. A. (1963). Introduction to the Theory of Statistics. McGraw-Hill, New York.

Palermo, D. S. and Jenkins, J. J. (1964). Word Association Norms. University of Minnesota Press, Minneapolis.

Pollio, H. R. (1963). A simple matrix analysis of association structure. J. Verb. Learn. and Verb. Beh. **2** 166-169.

Puff, C. R. (1964). Clustering as a function of the sequential organization of stimulus word lists. Unpublished Doctoral dissertation, University of Connecticut.

Rao, C. R. (1965). Linear Statistical Inference and Its Application. Wiley, New York.

Waugh, N. C. and Smith, J. E. K. (1962). A stochastic model for recall. Psychometrika. **27** 141-154.

STOCHASTIC MODELS IN A FREE-RECALL EXPERIMENT

by

DAVID LOYAL JENKINS

B. S., Fort Hays Kansas State College, 1964

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics and
Statistical Laboratory

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1967

ABSTRACT

The purpose of this paper is to show four stochastic models used to represent the data of a free recall experiment. The free-recall experiment for the first three models considered in this paper is one in which a subject is given as many trials as necessary to completely learn a list of words. In the last model considered the subject may only partially learn the list of words.

The first model considered is the Bush and Mosteller linear model. Changes in the probabilities of recall or non-recall are described with the aid of linear operators. By knowing only two parameters the data of a free-recall experiment can be summarized.

The next model considered is the Miller and McGill model. Their model is closely related to Bush and Mosteller's model. Using the estimates of Bush and Mosteller's model in Miller and McGill's model the probability of recalling a word exactly k times in n trials, and the expected number of times a word is recalled in n trials can be found.

The third model discussed is Waugh and Smith's stochastic model. The model describes a Markov process with a realizable absorbing state, allowing complete learning on some finite trial as well as imperfect retention prior to this trial.

The last model considered is the Cowan model. This model considers the effect of associations between words that will appear in a given recall position. The recall of words is regarded as a Markov chain where the category of the recalled word

is determined by the kind of word preceding it. Three parameters are used which are based on associative measures of between and within categories of stimulus words.